(4)

Report Number 1

Estimation of Parameters of Non-Gaussian Non-zero Mean Autoregressive Processes with Application to Optimal Detection in Colored Noise

Debasis Sengupta and Steven Kay

Department of Electrical Engineering
University of Rhode Island
Kingston, Rhode Island, 02881

June 1988

Prepared for

OFFICE OF NAVAL RESEARCH
Statistics and Probabilty Branch
Arlington, Virginia 22217
under Contract N00014-87-K-0021
S.M. Kay, Principal Investigator

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 1

2. GOVT ACCESSION NO.:

3. RECIPIENT'S CATALOG NUMBER:

4. TITLE (and Subtitle): Estimation of Parameters of Non-Caussian Non-zero Mean Autoregressive Processes with Application to Optimal Detection in Colored Noise

5. TYPE OF REPORT & PERIOD COVERED: Feb. 1987 to June 19'8

6. PERFORMING ORG. REPORT NUMBER:

7. AUTHOR(s):
Debasis Sengupta
Steven Kay

8. CONTRACT OR GRANT NUMBER(s): N00014-87-K-0021

9. PERFORMING ORGANIZATION NAME AND ADDRESS:
Electrical Engineering Department
University of Rhode Island
Kingston, RI 02881

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:

11. CONTROLLING OFFICE NAME AND ADDRESS:
Office of Naval Research
Department of the Navy
Arlington, Virginia 22217

12. REPORT DATE: June 1988

13. NUMBER OF PAGES: 36

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office):

15. SECURITY CLASS. (of this report): Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report):
Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report):

18. SUPPLEMENTARY NOTES:

19. KEY WORDS (Continue on reverse side if necessary and identify by block number):

Non-Gaussian Processes
Generalized Likelihood Ratio Test
Uniformly Most Powerful Test
Maximum Likelihood
Autoregressive Processes
Cramer-Rao bound

20. ABSTRACT (Continue on reverse side if necessary and identify by block number):
The problem addressed in this paper is that of estimating signal and noise parameters from a mixture of Non-Gaussian autoregressive (AR) noise with partially known deterministic signal. Two models are considered in order to examine different kinds of additive mixing. The Cramer-Rao bounds to the joint estimation of the signal amplitude and the noise parameters are presented. A computationally efficient estimator, which was previously proposed for estimation in the absence of signal, is extended for the two

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered) models under consideration. The proposed method essentially consists of two stages of least squares (LS) estimation which is motivated by the maximum likelihood estimation (MLE). The technique is then applied to the problem of detecting a signal known except for amplitude in colored non-Gaussian noise. Two slightly different mixing models are used and a generalized likelihood ratio test (GLRT), coupled with the proposed estimation scheme, is used to solve the problems. The results of computer simulations are presented as an evidence of the validity of the theoretical predictions of performance.

Accession For	
NTIS GRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

S N 0102- LF- 014- 6601

I. Introduction

The issue of estimating the parameters of non-Gaussian autoregressive (AR) processes has received significant consideration in recent years [Martin 1982], [Kay and Sengupta 1986, A]. These processes are modeled by an all-pole filter excited by zero-mean white noise, also called the *driving noise*, which may be non-Gaussian. In many applications, however, the driving noise can not be assumed to have zero mean. As an example, it may be necessary to process a deterministic signal contaminated by non-Gaussian noise after both the signal and the noise have passed through a channel. A method to estimate the parameters of a non-Gaussian non-zero mean AR process may be an useful tool for such processing. Another problem of interest is the estimation of the parameters of a zero-mean AR process when a combination of deterministic signals of unknown amplitude is added to it before it is observed. This paper suggests a generalization of an estimation technique proposed earlier [Kay and Sengupta 1987, Lee 1987] in order to make it applicable to both the non-zero mean processes described above. Essentially it is a two stage procedure based on an approximation of the maximum likelihood estimator (MLE). The resulting estimator is asymptotically efficient in the sense that its covariance matrix approaches the Fisher information matrix for large data records.

The paper is organized as follows. The Cramer-Rao bound for the parameters of each model is presented in Section II. Section III develops the two-stage least squares (LS) estimator for these parameters. In Section IV this method is applied to the problem of detecting a signal known except for amplitude in the presence of colored non-Gaussian noise using a generalized likelihood ratio test (GLRT). Although the GLRT is known to have nice asymptotic properties, it could not be used for this problem before due to the unavailability of a reasonably good estimation technique. Section V presents the results of computer simulation of the performance of the two-stage LS estimator. The performance of the GLRT detector for a typical detection problem is also simulated. Section VI summarizes the results.

II. Cramer-Rao Bounds

Consider N observations of the following non-zero mean AR(p) processes

$$x_n = -\sum_{j=1}^{p} a_j x_{n-j} + \nu^T \mathbf{s}_n + u_n \qquad (1)$$

$$x_n = \boldsymbol{\mu}^T \mathbf{s}_n - \sum_{j=1}^{p} a_j(x_{n-j} - \boldsymbol{\mu}^T \mathbf{s}_{n-j}) + u_n \tag{2}$$

where the driving noise u_n has a zero-mean PDF $f(u_n)$ with variance σ^2 and $\{\mathbf{s}_n\} = \{[s_n^{(1)}\ s_n^{(2)}\ \cdots\ s_n^{(q)}]^T\}$ is a known sequence of vectors of order q. $\boldsymbol{\nu}^T = [\nu_1\ \nu_2\ \cdots\ \nu_q]$ and $\boldsymbol{\mu}^T = [\mu_1\ \mu_2\ \cdots\ \mu_q]$ are the vectors of coefficients of $\mathbf{s}_n$ in models (1) and (2), respectively. f is assumed to be a symmetric function of u_n. It is also assumed that f has tails heavier than a Gaussian PDF having equal variance. This is characterized by the presence of impulses in the driving noise time series. (1) represents an AR process where a deterministic mean is added to the otherwise zero-mean driving noise (see Figure 1(a)). This model is also useful in some system identification problems [Eykhoff 1974]. (2) represents a zero-mean AR processes to which a deterministic signal has been added before it is observed (see Figure 1(b)). It is easy to observe that if $s_n^{(i)}$ is a geometric sequence

$$s_n^{(i)} = r_i^n \tag{3}$$

for $i = 1, 2, \cdots, q$, then (2) takes the form of (1) with

$$\nu_i = \mu_i \left(1 + \sum_{j=1}^{p} a_j r_i^{-j}\right) = \mu_i \sum_{j=0}^{p} a_j r_i^{-j}, \qquad i = 1, 2, \cdots, q \tag{4}$$

where a_0 is defined to be unity. This is not surprising, since (3) is the well-known eigenfunction of a linear shift invariant (LSI) filter. It can be shown that (3) is the only possible form of $s_n^{(i)}$ for which there is direct correspondence between the elements of $\boldsymbol{\nu}$ and $\boldsymbol{\mu}$. This special case includes real and complex exponentials, dc signals[Anderson 1971] and alternating signals. It can be generalized to the case of *pairwise* correspondence between the elements of $\boldsymbol{\nu}$ and $\boldsymbol{\mu}$ as follows. Consider for $q = 2$

$$\begin{aligned} s_n^{(1)} &= r^n \cos(n\omega) \\ s_n^{(2)} &= r^n \sin(n\omega) \end{aligned} \tag{5}$$

It follows that (2) is equivalent to (1) with

$$\begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix} = \begin{pmatrix} \sum_{j=0}^{p} r^{-j}\cos(j\omega) & -\sum_{j=0}^{p} r^{-j}\sin(j\omega) \\ \sum_{j=0}^{p} r^{-j}\sin(j\omega) & \sum_{j=0}^{p} r^{-j}\cos(j\omega) \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \tag{6}$$

This special case is useful in representing damped or undamped sinusoids with arbitrary phase. In general, $\mathbf{s}_n$ may be composed of a combination of functions of the form (3) and (5) in order to make (2) equivalent to (1). The importance of this equivalence will be apparent later.

The joint PDF of process (1) is given by

$$\mathbf{f_x}(\mathbf{x}) = \mathbf{f}_c(\mathbf{x})\mathbf{f}_p(\mathbf{x}) \tag{7}$$

where

$$\mathbf{f}_c(\mathbf{x}) = \prod_{n=p+1}^{N} f\left(\sum_{j=0}^{p} a_j x_{n-j} - \boldsymbol{\nu}^T \mathbf{s}_n\right)$$

$$\mathbf{f}_p(\mathbf{x}) = \mathbf{f}_{x_1\ x_2\ \cdots\ x_p}(x_1, x_2, \cdots, x_p)$$

The information matrix for the parameter vector $\boldsymbol{\Phi}$ can be shown to be

$$\begin{aligned} I_{\boldsymbol{\Phi}} &= -E\left[\frac{\partial^2 \ln \mathbf{f}_c(\mathbf{x})}{\partial \boldsymbol{\Phi} \partial \boldsymbol{\Phi}^T}\right] - E\left[\frac{\partial^2 \ln \mathbf{f}_p(\mathbf{x})}{\partial \boldsymbol{\Phi} \partial \boldsymbol{\Phi}^T}\right] \\ &= -\sum_{n=p+1}^{N} E\left[\frac{\partial^2 \ln f(u_n)}{\partial \boldsymbol{\Phi} \partial \boldsymbol{\Phi}^T}\right] - E\left[\frac{\partial^2 \ln \mathbf{f}_p(\mathbf{x})}{\partial \boldsymbol{\Phi} \partial \boldsymbol{\Phi}^T}\right] \end{aligned} \tag{8}$$

The second term accounts for the information from the first p samples while the first term corresponds to the information from all subsequent samples up to N. For large sample size the second term may be neglected. Defining the parameter vector as

$$\boldsymbol{\Phi} = [\boldsymbol{\nu}^T\ \mathbf{a}^T\ \sigma^2]^T \tag{9}$$

where $\mathbf{a} = [a_1\ a_2\ \cdots\ a_p]^T$, it is shown in the Appendix that

$$\mathbf{I}_{\boldsymbol{\Phi}} = \begin{pmatrix} I_f \mathbf{S}^T\mathbf{S} & \vdots & -I_f \mathbf{S}^T \mathbf{M} & \vdots & \mathbf{0}_q \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ -I_f \mathbf{M}^T \mathbf{S} & \vdots & I_f[(N-p)\mathbf{C} + \mathbf{M}^T\mathbf{M}] & \vdots & \mathbf{0}_p \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \mathbf{0}_q^T & \vdots & \mathbf{0}_p^T & \vdots & (N-p) I_{\sigma^2} \end{pmatrix} \tag{10}$$

where

$$I_f = E\left[\frac{\partial}{\partial u}\ln f(u)\right]^2$$

$$I_{\sigma^2} = \frac{1}{4\sigma^4}\left[E\left[u\frac{\partial}{\partial u}\ln f(u)\right]^2 - 1\right]$$

$$\mathbf{S} = \left[\mathbf{s}^{(1)}\ \mathbf{s}^{(2)}\ \cdots\ \mathbf{s}^{(q)}\right]$$

$$\mathbf{s}^{(i)} = \left[s_{p+1}^{(i)}\ s_{p+2}^{(i)}\ \cdots\ s_N^{(i)}\right]^T \qquad i = 1, 2, \cdots, q$$

$$\mathbf{M} = \begin{pmatrix} m_p & m_{p-1} & \cdots & m_1 \\ m_{p+1} & m_p & \cdots & m_2 \\ \vdots & \vdots & \ddots & \vdots \\ m_{N-1} & m_{N-2} & \cdots & m_{N-p} \end{pmatrix}$$

C is the $p \times p$ covariance matrix of the time series (1) and the sequence $\{m_1, m_2, \cdots, m_N\}$ is the mean of the time series (1). They satisfy the recursion

$$m_n = -\sum_{j=1}^{p} a_j m_{n-j} + \nu^T \mathbf{s}_n \tag{11}$$

which is the output of the filter excited by the signal $\{\nu^T \mathbf{s}_n\}$. The Cramer-Rao lower bounds for the parameters are obtained from the inverse of (10) (see Appendix).

$$Cov(\hat{\nu}) \geq \frac{1}{\sigma^2 I_f}\left[\mathbf{S}_n{}^T\left[\mathbf{I} + \mathbf{M}_n[(N-p)\mathbf{C}_n]^{-1}\mathbf{M}_n{}^T\right]^{-1}\mathbf{S}_n\right]^{-1} \tag{12a}$$

$$Cov(\hat{\mathbf{a}}) \geq \frac{1}{\sigma^2 I_f}\left[(N-p)\mathbf{C}_n + \mathbf{M}_n{}^T\left[\mathbf{I} - \mathbf{S}_n(\mathbf{S}_n{}^T\mathbf{S}_n)^{-1}\mathbf{S}_n{}^T\right]\mathbf{M}_n\right]^{-1} \tag{12b}$$

$$Var(\hat{\sigma}^2) \geq \frac{1}{(N-p)I_{\sigma^2}} \tag{12c}$$

The subscript n denotes normalization of the corresponding quantities by σ. The matrix inequalities (12a) and (12b) indicate that the difference of the left and right hand side matrices are nonnegative definite.

Since the matrix $\mathbf{M}_n[(N-p)\mathbf{C}_n]^{-1}\mathbf{M}_n{}^T$ is positive definite, the right hand side of (12a) approaches its minimum value of $[\sigma^2 I_f \mathbf{S}_n{}^T\mathbf{S}_n]^{-1}$ when $\mathbf{M}_n$ is close to zero. This requires the spectrums of the signals $\{s_n^{(i)}\}$, $i = 1, 2, \cdots, q$ to have a *mismatch* with the noise power spectral density (PSD). The lack of knowledge of the AR filter parameters always makes it more difficult to estimate ν. Note that the special case of *singular* $\mathbf{S}_n{}^T\mathbf{S}_n$ has been ignored. This can happen if and only if the columns of $\mathbf{S}_n$ are linearly dependent and consequently $\boldsymbol{\nu}$ is not entirely observable. A suitable reduction in the size of $\boldsymbol{\nu}$ is required to avoid this situation. The matrix $(\mathbf{I} - \mathbf{S}_n(\mathbf{S}_n^T\mathbf{S}_n)^{-1}\mathbf{S}_n^T)$ on the right hand side of $(12b)$ is idempotent. Therefore it is nonnegative definite and serves to reduce the CR bound by using the additional information about the AR filter parameters carried by the signal. The task of designing a suitable "probing" signal to extract a large amount of information about $\mathbf{a}$ is a problem of system identification [Eykhoff 1974]. Finally the quantity $\sigma^2 I_f$ is known to achieve its minimum value of unity only in the case of a Gaussian PDF [Sengupta and Kay 1986, A]. Hence $\boldsymbol{\nu}$ and $\mathbf{a}$ may be estimated more precisely in the case of a non-Gaussian PDF of the driving noise than in the Gaussian case. The quantity $\sigma^2 I_f$ depends on the shape of the PDF and is a quantitative measure of the expected improvement over the Gaussian case. To show the scale-invariance of $\sigma^2 I_f$ define g to be the PDF of the *normalized* random variable $\dot{u} = u/\sigma$,

$$g(\dot{u}) = \sigma f(u)$$

then

$$\begin{aligned}
\sigma^2 I_f &= \sigma^2 \int_{-\infty}^{\infty} \frac{[f'(u)]^2}{f(u)} du \\
&= \sigma^2 \int_{-\infty}^{\infty} \frac{\left[\frac{1}{\sigma} g'(\frac{u}{\sigma}) \frac{1}{\sigma}\right]^2}{\frac{1}{\sigma} g\left(\frac{u}{\sigma}\right)} du \\
&= \sigma^2 \int_{-\infty}^{\infty} \frac{\frac{1}{\sigma^4}(g'(\dot{u}))^2}{\frac{1}{\sigma} g(\dot{u})} \sigma d\dot{u} \\
&= \int_{-\infty}^{\infty} \frac{(g'(\dot{u}))^2}{g(\dot{u})} d\dot{u}
\end{aligned}$$

which does not depend on the variance of u. Hence $\sigma^2 I_f$ depends only on the *shape* of the PDF and is unaffected by scaling.

The joint PDF of process (2) also has the form (7) where

$$\mathbf{f}_c(\mathbf{x}) = \prod_{n=p+1}^{N} f\left(\sum_{j=0}^{p} a_j (x_{n-j} - \boldsymbol{\mu}^T \mathbf{s}_{n-j})\right)$$

The approximate information matrix (ignoring the contribution from the first p samples) for the parameter vector

$$\boldsymbol{\Psi} = [\boldsymbol{\mu}^T \ \mathbf{a}^T \ \sigma^2]^T \tag{13}$$

was derived by Martin [1982] for the special case $q = 1$ and $s_n^{(1)} = 1 \ \forall \ n$ and by the authors [Kay and Sengupta 1986, B] for the case $q = 1$, $s_n^{(1)}$ arbitrary. The generalization to the case $q > 1$ can be easily shown to be

$$\mathbf{I}_{\boldsymbol{\Psi}} = \begin{pmatrix} I_f \mathbf{V}^T\mathbf{V} & \vdots & \mathbf{0}_{p\times q}^T & \vdots & \mathbf{0}_q \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \mathbf{0}_{p\times q} & \vdots & (N-p)I_f\mathbf{C} & \vdots & \mathbf{0}_p \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \mathbf{0}_q^T & \vdots & \mathbf{0}_p^T & \vdots & (N-p)I_{\sigma^2} \end{pmatrix} \tag{14}$$

where

$$\begin{aligned} \mathbf{V} &= \left[\mathbf{v}^{(1)} \ \mathbf{v}^{(2)} \ \cdots \ \mathbf{v}^{(q)}\right] \\ \mathbf{v}^{(i)} &= \left[v_{p+1}^{(i)} \ v_{p+2}^{(i)} \ \cdots \ v_N^{(i)}\right]^T \qquad & i = 1, 2, \cdots, q \\ v_n^{(i)} &= \sum_{j=0}^{p} a_j s_{n-j}^{(i)}, \qquad & n = p+1, p+2, \cdots N \end{aligned}$$

The CR bounds are immediately obtained as

$$Cov(\hat{\boldsymbol{\mu}}) \geq \frac{1}{\sigma^2 I_f}(\mathbf{V}_n^T\mathbf{V}_n)^{-1} \tag{15a}$$

$$Cov(\hat{\mathbf{a}}) \geq \frac{1}{(N-p)\sigma^2 I_f}\mathbf{C}_n^{-1} \tag{15b}$$

$$Var(\hat{\sigma}^2) \geq \frac{1}{(N-p)I_{\sigma^2}} \tag{15c}$$

n once again denotes normalization. The lower bound on $Cov(\hat{\boldsymbol{\mu}})$ is small if $\mathbf{V}_n$ is large. This can be accomplished if there is a significant mismatch between the spectrums of the signal components $\{s_n^{(i)}\}$, $i = 1, 2, \cdots, q$ and the noise power spectral density (PSD) (see the definition of $\mathbf{V}$ in (14)). Higher signal amplitudes do not help reduce the CR bound. The lower bound on $Cov(\hat{\mathbf{a}})$ is independent of the shape of the signal. Both $Cov(\hat{\boldsymbol{\mu}})$ and $Cov(\hat{\mathbf{a}})$ have smaller lower bounds in the case of non-Gaussian PDF than in the Gaussian case and the difference is given by the factor $\sigma^2 I_f$ as before.

It is clear from (12b) and (15b) that the CR bound for $\hat{\mathbf{a}}$ is lower in the case of model (1) than that of model (2). It is now shown that if (3) holds then the bound (12b) attains its maximum value given by (15b). Substitution of (3) in (11) yields

$$\sum_{j=0}^{p} a_j m_{n-j} = \sum_{i=1}^{q} \nu_i r_i^n = \sum_{j=0}^{p} a_j \left(\sum_{i=1}^{q} \mu_i r_i^{n-j} \right)$$

i.e.,

$$m_n = \sum_{i=1}^{q} \mu_i r_i^n$$

Thus the kth column of $\mathbf{M}$ can be written as

$$\sum_{i=1}^{q} \frac{\mu_i}{r_i^k} \mathbf{s}^{(i)}$$

which belongs to the column space of S. Hence

$$\mathbf{M}^T[\mathbf{I} - \mathbf{S}(\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T]\mathbf{M} = \mathbf{0}$$

and consequently (12b) and (15b) are equivalent. This result has nice intuitive justification, since the models (1) and (2) become equivalent when (3) holds.

III. The MLE and its Approximation

From (5) the log-likelihood function for process (1) is given by

$$\ln \mathbf{f_x}(\mathbf{x}) = \sum_{n=p+1}^{N} \ln f \left(\sum_{j=0}^{p} a_j x_{n-j} - \boldsymbol{\nu}^T \mathbf{s}_n \right) + \ln \mathbf{f}_p(\mathbf{x})$$

The second term is ignored for the purpose of simplicity of maximization over the parameters. The remaining term is the well-known *conditional* likelihood function [Box and Jenkins 1970]. The MLE of $\boldsymbol{\Phi}$ is approximately a solution of

$$\frac{\partial}{\partial \boldsymbol{\Phi}} \sum_{n=p+1}^{N} \ln f \left(\sum_{j=0}^{p} a_j x_{n-j} - \boldsymbol{\nu}^T \mathbf{s}_n \right) = 0$$

Carrying out the differentiation with respect to ν_i,

$$\sum_{n=p+1}^{N} \frac{\partial}{\partial \nu_i} \ln f(u_n) \Big|_{u_n=\sum_{j=0}^{p} a_j x_{n-j} - \nu^T \mathbf{s}_n}$$
$$= \sum_{n=p+1}^{N} - s_n^{(i)} \frac{f'(u_n)}{f(u_n)} \Big|_{u_n=\sum_{j=0}^{p} a_j x_{n-j} - \nu^T \mathbf{s}_n}$$
$$= \sum_{n=p+1}^{N} s_n^{(i)} u_n \Gamma(u_n) \Big|_{u_n=\sum_{j=0}^{p} a_j x_{n-j} - \nu^T \mathbf{s}_n}$$
$$= 0$$

$$i.e., \quad \sum_{k=1}^{q} \nu_k \left(\sum_{n=p+1}^{N} s_n^{(i)} s_n^{(k)} \Gamma(u_n) \right) + \sum_{j=1}^{p} a_j \left(- \sum_{n=p+1}^{N} s_n^{(i)} x_{n-j} \right) \Gamma(u_n)$$
$$= \sum_{n=p+1}^{N} s_n^{(i)} x_n \Gamma(u_n) \quad i = 1, 2, \cdots, q \tag{16}$$

where

$$\Gamma(u) = -\frac{f'(u)}{u f(u)} \tag{17}$$

Differentiation with respect to a_j results

$$\sum_{n=p+1}^{N} \frac{\partial}{\partial a_j} \ln f(u_n) \Big|_{u_n=\sum_{j=0}^{p} a_j x_{n-j} - \nu^T \mathbf{s}_n} = \sum_{n=p+1}^{N} x_{n-j} \frac{f'(u_n)}{f(u_n)}$$
$$= - \sum_{n=p+1}^{N} x_{n-j} u_n \Gamma(u_n)$$
$$= 0, \qquad j = 1, 2, \cdots, p$$

$$i.e., \quad \sum_{k=1}^{q} \nu_k \left(- \sum_{n=p+1}^{N} s_n^{(k)} x_{n-j} \Gamma(u_n) \right) + \sum_{j=1}^{p} a_j \left(\sum_{n=p+1}^{N} x_{n-i} x_{n-j} \Gamma(u_n) \right)$$
$$= - \sum_{n=p+1}^{N} x_n x_{n-i} \Gamma(u_n), \qquad i = 1, 2, \cdots, p \tag{18}$$

Differentiation with respect to σ^2 requires one to define a scaled version of f

$$g(t) = \sigma f(\sigma t) \tag{19}$$

so that g has unity variance. It follows that

$$\sum_{n=p+1}^{N} \frac{\partial}{\partial \sigma^2} \ln f(u_n) \Big|_{u_n=\sum_{j=0}^{p} a_j x_{n-j} - \nu^T s_n}$$

$$= -\frac{1}{2\sigma^2} \sum_{n=p+1}^{N} \left[u_n \left(\frac{g'(\frac{u_n}{\sigma})}{\sigma g(\frac{u_n}{\sigma})} \right) + 1 \right] \Big|_{u_n=\sum_{j=0}^{p} a_j x_{n-j} - \nu^T s_n}$$

$$= -\frac{1}{2\sigma^2} \sum_{n=p+1}^{N} \left[u_n \frac{f'(u_n)}{f(u_n)} + 1 \right] \Big|_{u_n=\sum_{j=0}^{p} a_j x_{n-j} - \nu^T s_n}$$

$$= -\frac{1}{2\sigma^2} \left[(N-p) - \sum_{n=p+1}^{N} u_n^2 \Gamma(u_n) \right] \Big|_{u_n=\sum_{j=0}^{p} a_j x_{n-j} - \nu^T s_n}$$

$$= 0$$

$$i.e., \quad \frac{1}{N-p} \sum_{n=p+1}^{N} u_n^2 \Gamma(u_n) \Big|_{u_n=\sum_{j=0}^{p} a_j x_{n-j} - \nu^T s_n} = 1 \tag{20}$$

For most PDF's $\Gamma(u_n)$ is a complicated function of σ^2. If all the other parameters are known, σ^2 can be evaluated from (20) by a numerical search. In the case of a Gaussian PDF it has been shown that $\Gamma(u) \equiv 1/\sigma^2 \ \forall \ u$ [Kay and Sengupta 1986, A]. Therefore (20) reduces to the familiar form

$$\sigma^2 = \frac{1}{N-p} \sum_{n=p+1}^{N} \left(\sum_{j=0}^{p} a_j x_{n-j} - \nu^T s_n \right)^2 \tag{21}$$

which is just the averaged sum of the residuals. In this case (16) and (18) reduce to the linear equations

$$\sum_{k=1}^{q} \nu_k \left(\sum_{n=p+1}^{N} s_n^{(i)} s_n^{(k)} \right) + \sum_{j=1}^{p} a_j \left(- \sum_{n=p+1}^{N} s_n^{(i)} x_{n-j} \right) = \sum_{n=p+1}^{N} s_n^{(i)} x_n$$

$$i = 1, 2, \cdots, q$$

$$\sum_{k=1}^{q} \nu_k \left(- \sum_{n=p+1}^{N} s_n^{(k)} x_{n-j} \right) + \sum_{j=1}^{p} a_j \left(\sum_{n=p+1}^{N} x_{n-i} x_{n-j} \right) = - \sum_{n=p+1}^{N} x_n x_{n-i},$$

$$i = 1, 2, \cdots, p$$

In the matrix form

$$\begin{pmatrix} \mathbf{s}^{(1)^T}\mathbf{s}^{(1)} & \cdots & \mathbf{s}^{(1)^T}\mathbf{s}^{(q)} & \vdots & -\mathbf{s}^{(1)^T}\mathbf{x}_1 & \cdots & -\mathbf{s}^{(1)^T}\mathbf{x}_p \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{s}^{(q)^T}\mathbf{s}^{(1)} & \cdots & \mathbf{s}^{(q)^T}\mathbf{s}^{(q)} & \vdots & -\mathbf{s}^{(q)^T}\mathbf{x}_1 & \cdots & -\mathbf{s}^{(q)^T}\mathbf{x}_p \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ -\mathbf{x}_1^T\mathbf{s}^{(1)} & \cdots & -x_1^T\mathbf{s}^{(q)} & \vdots & \mathbf{x}_1^T\mathbf{x}_1 & \cdots & \mathbf{x}_1^T\mathbf{x}_p \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -\mathbf{x}_p^T\mathbf{s}^{(1)} & \cdots & -x_p^T\mathbf{s}^{(q)} & \vdots & \mathbf{x}_p^T\mathbf{x}_1 & \cdots & \mathbf{x}_p^T\mathbf{x}_p \end{pmatrix} \begin{pmatrix} \nu_1 \\ \vdots \\ \nu_q \\ \cdots \\ a_1 \\ \vdots \\ a_p \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{s}^{(1)^T}\mathbf{x}_0 \\ \vdots \\ \mathbf{s}^{(q)^T}\mathbf{x}_0 \\ \cdots \\ -\mathbf{x}_1^T\mathbf{x}_0 \\ \vdots \\ -\mathbf{x}_p^T\mathbf{x}_0 \end{pmatrix} \tag{22}$$

where

$$\mathbf{x}_j = [x_{p-j+1}\ x_{p-j+2}\ \cdots\ x_{N-j}]^T, \qquad j = 0, 1, \cdots, p$$

This is a generalization [Anderson 1971] of the *covariance method* of linear prediction which includes the estimation of $\boldsymbol{\nu}$ along with that of $\mathbf{a}$. The matrix on the left hand side can be easily shown to be positive definite with probability 1. (22) has a unique solution which is obtained in $\mathcal{O}(p^2)$ operations. The estimates of $\boldsymbol{\nu}$ and $\mathbf{a}$ can be used to find the MLE of σ^2 from (21).

In the general non-Gaussian case the equations (16) and (18) are highly nonlinear, since u_n is a linear function of $\boldsymbol{\nu}$ and $\mathbf{a}$ and Γ is usually a complicated function. The form of these equations resemble those of efficiency robust M-estimators [Huber 1981] for which Γ is chosen suitably to minimize the dependence of the efficiency of the estimator on the PDF [Martin 1979], [Lee 1987]. If, however, Γ is as defined in (17) and u_n is computed using some fixed and approximate value of $\boldsymbol{\nu}$ and $\mathbf{a}$, both the equations become linear. In the matrix form they can be written as

$$\underbrace{\begin{pmatrix} \mathbf{s}^{(1)^T}\mathbf{\Gamma}\mathbf{s}^{(1)} & \cdots & \mathbf{s}^{(1)^T}\mathbf{\Gamma}\mathbf{s}^{(q)} & \vdots & -\mathbf{s}^{(1)^T}\mathbf{\Gamma}\mathbf{x}_1 & \cdots & -\mathbf{s}^{(1)^T}\mathbf{\Gamma}\mathbf{x}_p \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{s}^{(q)^T}\mathbf{\Gamma}\mathbf{s}^{(1)} & \cdots & \mathbf{s}^{(q)^T}\mathbf{\Gamma}\mathbf{s}^{(q)} & \vdots & -\mathbf{s}^{(q)^T}\mathbf{\Gamma}\mathbf{x}_1 & \cdots & -\mathbf{s}^{(q)^T}\mathbf{\Gamma}\mathbf{x}_p \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ -\mathbf{x}_1^T\mathbf{\Gamma}\mathbf{s}^{(1)} & \cdots & -x_1^T\mathbf{\Gamma}\mathbf{s}^{(q)} & \vdots & \mathbf{x}_1^T\mathbf{\Gamma}\mathbf{x}_1 & \cdots & \mathbf{x}_1^T\mathbf{\Gamma}\mathbf{x}_p \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -\mathbf{x}_p^T\mathbf{\Gamma}\mathbf{s}^{(1)} & \cdots & -x_p^T\mathbf{\Gamma}\mathbf{s}^{(q)} & \vdots & \mathbf{x}_p^T\mathbf{\Gamma}\mathbf{x}_1 & \cdots & \mathbf{x}_p^T\mathbf{\Gamma}\mathbf{x}_p \end{pmatrix}}_{\mathbf{X}} \begin{pmatrix} \nu_1 \\ \vdots \\ \nu_q \\ \cdots \\ a_1 \\ \vdots \\ a_p \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{s}^{(1)^T}\mathbf{\Gamma}\mathbf{x}_0 \\ \vdots \\ \mathbf{s}^{(q)^T}\mathbf{\Gamma}\mathbf{x}_0 \\ \cdots \\ -\mathbf{x}_1^T\mathbf{\Gamma}\mathbf{x}_0 \\ \vdots \\ -\mathbf{x}_p^T\mathbf{\Gamma}\mathbf{x}_0 \end{pmatrix} \tag{23}$$

where $\mathbf{\Gamma} = diag\{\Gamma(\hat{u}_{p+1}), \Gamma(\hat{u}_{p+2}), \cdots, \Gamma(\hat{u}_N)\}$ and

$$\hat{u}_n = x_n + \sum_{j=1}^{p} \hat{a}_j x_{n-j} - \hat{\boldsymbol{\nu}}^T \mathbf{s}_n, \qquad n = p+1, p+2, \cdots, N \tag{24}$$

which uses fixed and approximate values of $\boldsymbol{\nu}$ and $\mathbf{a}$ obtained from an initial stage of unweighted LS given by (22). (23) can be interpreted as the solution to the ***weighted*** least squares (LS) problem

$$\min_{\boldsymbol{\nu},\mathbf{a}} \sum_{n=p+1}^{N} \left(x_n + \sum_{j=1}^{p} a_j x_{n-j} - \boldsymbol{\nu}^T \mathbf{s}_n \right)^2 \Gamma(\hat{u}_n) \tag{25}$$

The solution to (23) is expected to be much better than the unweighted LS estimators, such as the covariance estimator, which implicitly assumes the underlying PDF to be Gaussian. $\mathbf{X}$ is a symmetric $(p+1) \times (p+1)$ matrix which is positive definite with probability 1. To show this write $\mathbf{X}$ as

$$\mathbf{X} = [\mathbf{s}^{(1)}\ \mathbf{s}^{(2)}\ \cdots\ \mathbf{s}^{(q)}\ \mathbf{x}_1\ \mathbf{x}_2\ \cdots\ \mathbf{x}_p]^T \mathbf{\Gamma} [\mathbf{s}^{(1)}\ \mathbf{s}^{(2)}\ \cdots\ \mathbf{s}^{(q)}\ \mathbf{x}_1\ \mathbf{x}_2\ \cdots\ \mathbf{x}_p]$$

$\Gamma(u)$ is positive for all the symmetric PDF's which are monotonically nonincreasing functions of positive values of u. This mild condition makes sure that the matrix $\mathbf{\Gamma}$ has positive diagonal elements. If $\mathbf{b} = [b_0\ b_1\ \cdots\ b_{p+q}]^T$ is a vector of real numbers then $\mathbf{b}^T\mathbf{X}\mathbf{b}$ is the squared norm of the random vector

$$\mathbf{\Gamma}^{1/2}[\mathbf{s}^{(1)}\ \mathbf{s}^{(2)}\ \cdots\ \mathbf{s}^{(q)}\ \mathbf{x}_1\ \mathbf{x}_2\ \cdots\ \mathbf{x}_p]\mathbf{b}$$

and must be greater than zero with probability 1. Since $\mathbf{X}$ is positive definite, (25) has a unique solution. Further simplification can be made by approximating Γ by a simpler function whenever it is complicated [Kay and Sengupta 1986, A]. σ^2 can be estimated from (20) after $\boldsymbol{\nu}$ and a are determined. Alternatively, a coarse estimate of σ^2 can be obtained by substituting $\hat{\boldsymbol{\nu}}$ and $\hat{\mathbf{a}}$ in (21). The suggested method is an extension of the two-stage LS estimator proposed in an earlier paper [Kay and Sengupta 1987].

The log-likelihood function for process (2) is given by

$$\ln \mathbf{f_x}(\mathbf{x}) = \sum_{n=p+1}^{N} \ln f\left(\sum_{j=0}^{p} a_j(x_{n-j} - \boldsymbol{\mu}^T\mathbf{s}_{n-j})\right) + \ln \mathbf{f}_p(\mathbf{x})$$

The second term is ignored, as before, leaving the conditional likelihood function. The MLE of $\boldsymbol{\Psi}$ is approximately a solution of

$$\frac{\partial}{\partial \boldsymbol{\Psi}} \sum_{n=p+1}^{N} \ln f\left(\sum_{j=0}^{p} a_j(x_{n-j} - \boldsymbol{\mu}^T\mathbf{s}_{n-j})\right) = 0$$

Differentiation with respect to σ^2 yields an equation similar to (20)

$$\frac{1}{N-p}\sum_{n=p+1}^{N} u_n^2\Gamma(u_n)\Bigg|_{u_n=\sum_{j=0}^{p} a_j(x_{n-j}-\boldsymbol{\mu}^T\mathbf{s}_{n-j})} = 1 \qquad (26)$$

Differentiation with respect to $\boldsymbol{\mu}$ and a, however, does not give rise to convenient forms like (16) and (18). In the special cases of (3) and (5) this process reduces to process (1). With the reparameterizations (4) and (6) one can proceed to estimate $\boldsymbol{\nu}$ and a using the estimation scheme discussed above. $\hat{\boldsymbol{\mu}}$ can then be computed from the inverse mapping of (4) and (6). Finally, $\hat{\sigma}^2$ is obtained by substituting estimates of $\boldsymbol{\mu}$ and a in (26).

Thus the two-stage LS estimator offers an elegant method of estimating the AR filter parameters and the driving noise variance along with the signal parameter ν for model (1). Any combination of signal shapes can be used as long as they are linearly independent. In the case of model (2) the applicability of the proposed technique is restricted to special signal shapes. This underscores the significance of the shapes described by (3) and (5), which include finite combination of sinusoids, exponentials and dc signals. The following section illustrates how the two-stage LS estimator can be used for the detection of such useful signals in nonwhite non-Gaussian noise.

IV. Application to Optimal Detection

Two detection problems will be discussed in this section. The first problem is to detect whether a signal is present when the noise as well as the signal, if any, has passed through a channel. If the unknown channel is modeled by an AR filter, the detection problem can be formulated as a test of the following hypotheses.

$$\begin{aligned} \mathcal{H}_0 : \quad x_n &= -\sum_{j=1}^{p} a_j x_{n-j} + u_n \\ \mathcal{H}_1 : \quad x_n &= -\sum_{j=1}^{p} a_j x_{n-j} + u_n + \nu^T \mathbf{s}_n \end{aligned} \tag{27}$$

(see Figure 1(a)). The noise u_n is assumed to have an even PDF known except for its variance σ^2. The sequence of vector signals $\{\mathbf{s}_n\}$ is assumed to be known except for its amplitude ν.

The second problem is to decide whether a signal is present in the presence of colored ambient noise [Kay and Sengupta 1986, B]. The signal shape is assumed to be known at the point of observation irrespective of the PSD of the noise. This is in contrast to the first problem where the signal goes through the same unknown process of colorization as the noise before it is observed. If the noise PSD is described by an AR model, the second problem is to choose between the hypotheses

$$\begin{aligned} \mathcal{H}_0 : \quad x_n &= -\sum_{j=1}^{p} a_j x_{n-j} + u_n \\ \mathcal{H}_1 : \quad x_n &= \mu^T \mathbf{s}_n - \sum_{j=1}^{p} a_j (x_{n-j} - \mu^T \mathbf{s}_{n-j}) + u_n \end{aligned} \tag{28}$$

(see Figure 1(b)). The same assumptions about the driving noise are made. The signal is assumed to be known except for its amplitude $\boldsymbol{\mu}$. Note that the time series equations for the two problems under $\mathcal{H}_1$ are identical to (1) and (2), respectively.

(27) and (28) together represent a large class of problems. The unknown set of parameters a allows for the unknown correlation pattern of the noise. The PDF f of u_n can be chosen to characterize specific problems in a realistic way. Finally, by allowing the amplitude of the signal to be unknown, the detector is expected to be tolerant to different attenuations and phase changes of different components of the signal.

The above two problems can be recast as

$$\begin{aligned} \mathcal{H}_0 &: \boldsymbol{\theta}^T = \left[\mathbf{0}^T \ \boldsymbol{\theta}_t^T\right] \\ \mathcal{H}_1 &: \boldsymbol{\theta}^T = \left[\boldsymbol{\theta}_q^T \ \boldsymbol{\theta}_t^T\right] \qquad \boldsymbol{\theta}_q \neq \mathbf{0} \end{aligned} \tag{29}$$

where $\boldsymbol{\theta}_t = [\mathbf{a}^T \ \sigma^2]^T$, $\boldsymbol{\theta}_q = \boldsymbol{\nu}$ in problem (27) and $\boldsymbol{\theta}_q = \boldsymbol{\mu}$ in problem (28). In either case the dimensions of $\boldsymbol{\theta}_t$ and $\boldsymbol{\theta}_q$ are $t = p+1$ and q, respectively. The two problems will be discussed side by side because of their close resemblence. It is well known that there is no uniformly most powerful (UMP) test for (29). Yet the generalized likelihood ratio test (GLRT) is widely preferred because of its nice asymptotic (large sample size) properties such as consistency, unbiasedness and constant false alarm rate (CFAR). It is also called the uniformly most powerful invariant (UMPI) test since it exhibits the UMP property among the class of tests which are invariant to a natural set of transformations [Lehmann 1959]. The asymptotic performance of the GLRT becomes equivalent to that of a clairvoyant GLRT (*i.e.*, the test with perfectly known values of $\boldsymbol{\theta}_t$) under the condition [Kendall and Stuart 1979]

$$\mathbf{I}_{\boldsymbol{\theta}_q\boldsymbol{\theta}_t}(\mathbf{0}, \boldsymbol{\theta}_t) = E\left[\left(\frac{\partial \ln \mathrm{f}_{\mathbf{x}}(\mathbf{x}; \boldsymbol{\theta}_q, \boldsymbol{\theta}_t)}{\partial \boldsymbol{\theta}_q}\right)\left(\frac{\partial \ln \mathrm{f}_{\mathbf{x}}(\mathbf{x}; \boldsymbol{\theta}_q, \boldsymbol{\theta}_t)}{\partial \boldsymbol{\theta}_t}\right)^T\right]\Bigg|_{\boldsymbol{\theta}_q = \mathbf{0}} = \mathbf{0} \tag{30}$$

$\mathbf{I}_{\boldsymbol{\theta}_q\boldsymbol{\theta}_t}(\boldsymbol{\theta}_q, \boldsymbol{\theta}_t)$ is identically zero for problem (28) (see (14)). From (10) it is apparent that (30) holds for problem (27) as well. Hence for both the problems the GLRT can be said to be asymptotically optimal in the sense that for large sample size the knowledge of a and σ^2, the so-called *nuisance parameters*, would not improve the performance.

The GLRT for testing (29) is to decide $\mathcal{H}_1$ if

$$\ell_G = \frac{\mathcal{L}(\hat{\boldsymbol{\theta}}_q, \hat{\boldsymbol{\theta}}_t)}{\mathcal{L}(\mathbf{0}, \hat{\hat{\boldsymbol{\theta}}}_t)} > \gamma \tag{31}$$

for some threshold γ, where $\mathcal{L}$ is the likelihood function

$$\mathcal{L}(\boldsymbol{\theta}_q, \boldsymbol{\theta}_t) = \mathbf{f}_{\mathbf{x}}(\mathbf{x}; \boldsymbol{\theta}_q, \boldsymbol{\theta}_t)$$

$\hat{\boldsymbol{\theta}}_t$ is the MLE of $\boldsymbol{\theta}_t$ assuming $\mathcal{H}_0$ is true while $\hat{\hat{\boldsymbol{\theta}}}_q$ and $\hat{\hat{\boldsymbol{\theta}}}_t$ are *joint* MLE's of $\boldsymbol{\theta}_q$ and $\boldsymbol{\theta}_t$ assuming $\mathcal{H}_1$ is true. $\hat{\boldsymbol{\theta}}_t$ is found by maximizing $\mathcal{L}(\mathbf{0}, \boldsymbol{\theta}_t)$ over $\boldsymbol{\theta}_t$. Similarly, $\hat{\hat{\boldsymbol{\theta}}}_q$, $\hat{\hat{\boldsymbol{\theta}}}_t$ are obtained by maximizing $\mathcal{L}(\boldsymbol{\theta}_q, \boldsymbol{\theta}_t)$ over $\boldsymbol{\theta}_q$ and $\boldsymbol{\theta}_t$.

The statistics of ℓ_G, the *likelihood ratio*, are difficult to obtain in general. For large data records (asymptotically) it may be shown that $2\ln \ell_G$ is distributed in the following manner [Kendall and Stuart 1979].

$$2\ln \ell_G \sim \chi_q^2 \qquad \text{under } \mathcal{H}_0 \tag{32a}$$

$$2\ln \ell_G \sim \chi'^2(q, \lambda) \qquad \text{under } \mathcal{H}_1 \tag{32b}$$

Here χ_q^2 represents a chi-square distribution with q degrees of freedom and $\chi'^2(q,\lambda)$ represents a noncentral chi-square distribution with q degrees of freedom and noncentrality parameter λ. Note that $\chi'^2(q,0) = \chi_q^2$ or the distribution under $\mathcal{H}_0$ is a special case of the distribution under $\mathcal{H}_1$ and occurs when $\lambda = 0$. If (30) holds, the noncentrality parameter λ, which is a measure of the discrimination between the two hypotheses, is given by

$$\lambda = \boldsymbol{\theta}_q^T \left[\mathbf{I}_{\boldsymbol{\theta}_q \boldsymbol{\theta}_q}(\mathbf{0}, \boldsymbol{\theta}_t)\right] \boldsymbol{\theta}_q = \boldsymbol{\theta}_q^T E\left[\left(\frac{\partial \ln \mathbf{f}_{\mathbf{x}}(\mathbf{x}; \boldsymbol{\theta}_q, \boldsymbol{\theta}_t}{\partial \boldsymbol{\theta}_q}\right)\left(\frac{\partial \ln \mathbf{f}_{\mathbf{x}}(\mathbf{x}; \boldsymbol{\theta}_q, \boldsymbol{\theta}_t}{\partial \boldsymbol{\theta}_q}\right)^T\right]\Bigg|_{\boldsymbol{\theta}_q = \mathbf{0}} \boldsymbol{\theta}_q \tag{33}$$

where $\boldsymbol{\theta}_q$, $\boldsymbol{\theta}_t$ are the true values. The probability of deciding $\mathcal{H}_1$ when $\mathcal{H}_0$ is true (also called the probability of *false alarm*) is given by

$$P_{FA} = \Pr\{2\ln \ell_G > \gamma' | \mathcal{H}_0\} \tag{34a}$$

where $\gamma' = 2\ln\gamma$. The probability of correctly deciding $\mathcal{H}_1$ (called the *power* of the test) is

$$P_D = \Pr\{2\ln \ell_G > \gamma' | \mathcal{H}_1\} \tag{34b}$$

In practice, γ' can be set to produce a given false alarm rate and P_D can be calculated from (34b) accordingly.

The likelihood ratio for problems (27) and (28) have the common form (see (7))

$$\ell_G = \frac{\mathrm{f_x}(\mathrm{x};\hat{\hat{\theta}}_q,\hat{\hat{\theta}}_t)}{\mathrm{f_x}(\mathrm{x};0,\hat{\theta}_t)} = \frac{\mathrm{f}_c(\mathrm{x};\hat{\hat{\theta}}_q,\hat{\hat{\theta}}_t)}{\mathrm{f}_c(\mathrm{x};0,\hat{\theta}_t)}\cdot\frac{\mathrm{f}_p(\mathrm{x};\hat{\hat{\theta}}_q,\hat{\hat{\theta}}_t)}{\mathrm{f}_p(\mathrm{x};0,\hat{\theta}_t)}$$

The second factor is dropped for ease of computation. A heuristic justification for ignoring the second term is that its contribution to ℓ_G will be negligible when N is large and the two hypotheses are close to each other. With this simplification, the test is equivalent to deciding $\mathcal{H}_1$ if

$$2\ln\ell_G = \ln\mathrm{f}_c(\mathrm{x};\hat{\hat{\theta}}_q,\hat{\hat{\theta}}_t) - \ln\mathrm{f}_c(\mathrm{x};0,\hat{\theta}_t) > \gamma' \tag{35}$$

Specifically for problem (27),

$$\begin{aligned}2\ln\ell_G = &\sum_{n=p+1}^{N}\ln f\left(\sum_{j=0}^{p}\hat{\hat{\mathrm{a}}}_j x_{n-j} - \hat{\hat{\nu}}^T \mathrm{s}_n;\hat{\hat{\sigma}}^2\right)\\ &-\sum_{n=p+1}^{N}\ln f\left(\sum_{j=0}^{p}\hat{\mathrm{a}}_j x_{n-j};\hat{\sigma}^2\right)\end{aligned} \tag{36a}$$

and for problem (28),

$$\begin{aligned}2\ln\ell_G = &\sum_{n=p+1}^{N}\ln f\left(\sum_{j=0}^{p}\hat{\hat{\mathrm{a}}}_j (x_{n-j} - \hat{\hat{\mu}}^T \mathrm{s}_{n-j});\hat{\hat{\sigma}}^2\right)\\ &-\sum_{n=p+1}^{N}\ln f\left(\sum_{j=0}^{p}\hat{\mathrm{a}}_j x_{n-j};\hat{\sigma}^2\right)\end{aligned} \tag{36b}$$

Hats and double hats once again indicate MLE's under $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. $\hat{\hat{\mathrm{a}}}_0$ and $\hat{\mathrm{a}}_0$ are both defined to be unity. Figures 2(a) and 2(b) are block diagrams of (36a) and (36b), respectively. In the Gaussian case $\ln f$ is a simple quadratic and the above test Statistics reduces to

$$2\ln\ell_G = (N-p)\ln\left(\frac{\hat{\sigma}^2}{\hat{\hat{\sigma}}^2}\right) \tag{37a}$$

with

$$\hat{\sigma}^2 = \frac{1}{N-p}\sum_{n=p+1}^{N}\left(\sum_{j=0}^{p}\hat{\mathrm{a}}_j x_{n-j}\right)^2$$

$$\hat{\hat{\sigma}}^2 = \frac{1}{N-p}\sum_{n=p+1}^{N}\left(\sum_{j=0}^{p}\hat{\hat{\mathrm{a}}}_j x_{n-j} - \hat{\hat{\nu}}^T \mathrm{s}_n\right)^2$$

for problem (27) and

$$2 \ln \ell_G = (N-p) \ln \left(\frac{\hat{\sigma}^2}{\hat{\hat{\sigma}}^2} \right) \tag{37b}$$

with

$$\hat{\sigma}^2 = \frac{1}{N-p} \sum_{n=p+1}^{N} \left(\sum_{j=0}^{p} \hat{a}_j x_{n-j} \right)^2$$

$$\hat{\hat{\sigma}}^2 = \frac{1}{N-p} \sum_{n=p+1}^{N} \left(\sum_{j=0}^{p} (\hat{a}_j x_{n-j} - \hat{\boldsymbol{\mu}}^T \mathbf{s}_{n-j}) \right)^2$$

for problem (28).

In order to compute the probabilities of false alarm and detection from (34), it is necessary to evaluate the noncentrality parameter. For problem (27) it is obtained from (33) and (10) as

$$\lambda = \boldsymbol{\nu}^T \mathbf{S}^T \mathbf{S} \boldsymbol{\nu} I_f = \frac{1}{\sigma^2} \mathbf{s}_0^T \mathbf{s}_0 [\sigma^2 I_f] \tag{38}$$

where $\mathbf{s}_0 = \mathbf{S}\boldsymbol{\nu}$ is the vector of signal sequence including the amplitude. The noncentrality parameter for problem (28) is obtained from (33) and (14).

$$\lambda = \boldsymbol{\mu}^T \mathbf{V}^T \mathbf{V} \boldsymbol{\mu} I_f \approx \frac{1}{\sigma^2} \boldsymbol{\mu}^T (\mathbf{AS})^T (\mathbf{AS}) \boldsymbol{\mu} [\sigma^2 I_f]$$

where $\mathbf{A}$ is the $(N-p) \times (N-p)$ Toeplitz matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ a_1 & \ddots & \ddots & & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & & \vdots \\ a_p & & \ddots & \ddots & \ddots & & \vdots \\ 0 & \ddots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & & \ddots & \ddots & 0 \\ 0 & \ddots & 0 & a_p & \cdots & a_1 & 1 \end{pmatrix}$$

Therefore

$$\lambda = \frac{1}{\sigma^2} \boldsymbol{\mu}^T \mathbf{S}^T (\mathbf{A}^T \mathbf{A}) \mathbf{S} \boldsymbol{\mu} [\sigma^2 I_f] = \mathbf{s}_0^T \mathbf{R}^{-1} \mathbf{s}_0 [\sigma^2 I_f] \tag{39}$$

where $\mathbf{s}_0 = \mathbf{S}\boldsymbol{\mu}$ and $\mathbf{R} = \sigma^2 (\mathbf{A}^T \mathbf{A})^{-1}$ is approximately the $(N-p) \times (N-p)$ covariance matrix of the noise. $\mathbf{s}_0^T \mathbf{R}^{-1} \mathbf{s}_0$ is the signal to noise ratio (SNR) at the output of a prewhitener built with perfect knowledge of the filter parameter a. To be more precise, if the data is passed through an ideal whitener (a filter which will completely whiten the *noise*), then $\mathbf{s}_0^T\mathbf{R}^{-1}\mathbf{s}_0$ is the ratio of the energy contributions from the signal and noise parts of the whitener output. The same interpretation holds for $\mathbf{s}_0^T\mathbf{s}_0/\sigma^2$ in (38) for the problem (27). In either case λ is proportional to the SNR at the output of a perfect prewhitener and correlator. From (39) it was shown earlier [Kay and Sengupta 1986, B] that the probability of detection, which is a monotonically increasing function of λ, can be improved in the case of problem (28) by choosing the signal to be one sinusoid along the direction of the weakest noise eigenvector, if any such knowledge is available at all. On the other hand the performance would deteriorate if the spectrum of the signal ($\mathbf{s}_0$) matches the noise PSD. From (38) it is clear, however, that problem (27) offers no such flexibility. This makes intuitive sense, since both the signal and the noise pass through the same filter. The factor $\sigma^2 I_f$ indicates the expected amount of reduction in the SNR required to achieve a given probability of detection for a non-Gaussian noise PDF over a Gaussian PDF.

In order to implement the GLRT detectors (36*a*) and (36*b*) it is necessary to compute the MLEs of the unknown parameters under each hypothesis. Alternatively, the two-stage LS estimator described in the previous section could be used as an approximation to the MLEs. The simpler version of the estimator for estimation under $\mathcal{H}_0$ was presented in an earlier paper [Kay and Sengupta 1987]. The extension provided in this paper allows one to choose any signal shape for (27), while only sinusoids, exponentials or dc signals, which satisfy (3) or (5), can be used in the case of problem (28). Computationally simpler approximations of the GLRT such as the Rao efficient score test [Rao 1973] can be used for more general signal shapes in (28). Under the asssumption of large sample size and weak signal the Rao statistic ℓ_R is equivalent to $2\ln \ell_G$. When (30) holds, the Rao statistic for testing (29) is

$$\ell_R = \left[\frac{\partial}{\partial \boldsymbol{\theta}_q} \ln \mathcal{L}(0, \hat{\boldsymbol{\theta}}_t)\right]^T \mathbf{I}^{-1}_{\boldsymbol{\theta}_q \boldsymbol{\theta}_q}(0, \hat{\boldsymbol{\theta}}_t) \left[\frac{\partial}{\partial \boldsymbol{\theta}_q} \ln \mathcal{L}(0, \hat{\boldsymbol{\theta}}_t)\right]$$

The Rao statistic uses MLEs under the null hypothesis only. In the case of problem (27) ℓ_R can be shown to be

$$\ell_R = \frac{1}{I_f}\mathbf{h}^T\mathbf{S}\left[\mathbf{S}^T\mathbf{S}\right]^{-1}\mathbf{S}^T\mathbf{h} \tag{40}$$

where

$$\mathbf{h} = [f'(\hat{u}_{p+1})/f(\hat{u}_{p+1}) \; f'(\hat{u}_{p+2})/f(\hat{u}_{p+2}) \; \cdots \; f'(\hat{u}_N)/f(\hat{u}_N)]^T$$

$$\mathrm{u}_n = \sum_{j=0}^{p} \hat{a}_j x_{n-j}, \qquad n = p+1, p+2, \cdots, N$$

In the case of problem (28) ℓ_R reduces to

$$\ell_R = \frac{1}{I_f} \mathbf{h}^T \mathbf{V} \left[\mathbf{V}^T \mathbf{V}\right]^{-1} \mathbf{V}^T \mathbf{h} \tag{41}$$

with $\mathbf{h}$ and $\hat{u}_n$ as defined above. In general $\mathbf{V}$ will be a function of $\hat{\mathbf{a}}$. In the special case of (3) it is observed that $\mathbf{V} = \mathbf{SD}$ with

$$\mathbf{D} = diag\left(\sum_{j=0}^{p} \hat{a}_j r_1^{-j}, \sum_{j=0}^{p} \hat{a}_j r_2^{-j}, \cdots, \sum_{j=0}^{p} \hat{a}_j r_q^{-j}\right)$$

Consequently (41) is equivalent to (40). The equivalence holds by a similar argument in the special case of (5) and in general for all signals which are composed of exponentials, damped and undamped sinusoids and dc. The two-stage LS estimator of a in the zero-mean case [Kay and Sengupta 1987] may be used in (40) and (41) instead of its MLE. Exclusion of the MLEs under $\mathcal{H}_1$ makes the Rao detector attractive from the estimation point of view. However its performance is expected to be worse than that of the GLRT detector for short sample sizes and large signal amplitudes, since the former involves more optimistic approximations.

V. Simulation of Performance

Two non-zero mean AR(4) processes of the form (2) were chosen for computer simulations. The AR filter parameters for the processes [Kay and Sengupta 1986,A] are given in Table A. Process I is broadband while process II is narrowband. A single signal component ($q = 1$) with amplitude $\mu = 0.1$ was used. The known part of the signal was chosen to be

$$s_n = (-1)^n \tag{42}$$

for $n = 1, 2, \cdots, N$. This is of the form (3) and therefore the process can be described by (1) as well. The equivalent values of ν corresponding to the processes

I and II are 0.4592 and 1.1147, respectively. The driving noise was assumed to be mixed-Gaussian with PDF

$$f(u) = (1-\epsilon)G_B(u) + \epsilon G_I(u) \tag{43}$$

where $G_B(u)$ and $G_I(u)$ are zero-mean Gaussian PDFs with variance σ_B^2 and σ_I^2, respectively. The mixture parameter ϵ is a number between 0 and 1 and determines the composition of the PDF. With the assumption $\sigma_I^2 >> \sigma_B^2$, (43) represents a nominally Gaussian background distributiuon G_B contaminated by an interfering component G_I with higher variance. This model can represent the background noise in radar and sonar communication where clutter and reverberation gives rise to occasional impulses in an otherwise Gaussian ambient noise. The variance of the overall PDF is

$$\sigma^2 = (1-\epsilon)\sigma_B^2 + \epsilon\sigma_I^2 \tag{44}$$

Simulations were carried out with $\sigma_B^2 = 1$, $\sigma_I^2 = 100$ and $\epsilon = 0.1$. The resulting variance is $\sigma^2 = 10.9$. The AR process was generated by passing a white mixed-Gaussian process through a filter, allowing sufficient time for the transients to decay. The white process was generated by randomly selecting from two mutually independent zero-mean white Gaussian processes with PDFs G_B and G_I on the basis of a series of Bernoulli trials with probability of success ϵ. Thus a random variable could be expected to come from the background population for $(1-\epsilon)$ fraction of the time and from the contaminating population for ϵ fraction of the time. σ_B^2 and σ_I^2 were assumed to be known while ϵ was assumed unknown. ϵ is linearly related to σ^2 by (44). Estimating ϵ is therefore equivalent to estimating σ^2.

Table B shows the sample means and sample variances of the unweighted LS estimators of a, μ, ν and σ^2 obtained by solving (22), (4) and (21). The performance is poor, as expected. The variance of the LS estimators of a, μ and ν are larger than the corresponding CR bound by roughly a factor of 10. This confirms the well-known result that the asymptotic variance of these LS estimators are given by the "Gaussian" CR bound irrespective of the driving noise PDF [Anderson 1971]. The "Gaussian" CR bound, which assumes the PDF to be Gaussian, is described by (15*b*), (15*a*) and (12*a*) with $\sigma^2 I_f = 1$ and is clearly larger than the true CR bound by a factor of $\sigma^2 I_f$. This factor, which is approximately 10 for the selected PDF parameters, explains the discrepancy between the last two columns of Table B. The results are based on 500 experiments, each conducted with 500 data points.

Table C summarizes the performance of the two-stage LS estimator obtained by solving (23), (4) and (26). The variances of the estimators of a, μ and ν are close to the CR bound. This improvement of performance is in accordance with similar results reported by the authors in the zero-mean case [Kay and Sengupta, 1987]. The variance of the estimator of σ^2 continues to be much larger than the CR bound. This is to be expected, since the estimator given by (21) is not an approximation of the MLE in any sense, the property of asymptotic efficiency of the MLE being inapplicable. The bias of all the estimators in Table C appear to be significantly less than the bias of their counterparts in Table B.

The same AR(4) processes were used for simulations of the performance of the GLRT and Rao detectors. A single signal component of the form of (42) was chosen. μ was adjusted to yield different values of SNR in the following way. The SNR as defined in the previous section is

$$\mathrm{SNR} = \mathbf{s}_0^T \mathbf{s}_0 = \nu^2 \sum_{n=p+1}^{N} s_n^2 = (N-p)\nu^2$$

Using (4),

$$\mathrm{SNR} = \frac{(N-4)\mu^2}{1 - a_1 + a_2 - a_3 + a_4}.$$

Therefore

$$\mu = \sqrt{\frac{(1 - a_1 + a_2 - a_3 + a_4)\mathrm{SNR}}{N-4}} \tag{45}$$

Thus μ was calculated for a given process such as to produce a desired SNR. The number of data points (N) was chosen to be 50. A probability of false alarm $P_{FA} = 0.01$ was used to evaluate the detection performance. The value of γ' necessary for this is 6.635, as obtained from the tables of central χ^2 distribution with one degree of freedom. The asymptotic performance of the GLRT or Rao detector is described by (32b) and (34b). The noncentrality parameter, as obtained from (38) or (39) is

$$\lambda = \mathrm{SNR}[\sigma^2 I_f]$$

λ is calculated for each value of the SNR and the theoretical or asymptotic performance is evaluated accordingly.

Since the signal has only one component, the Rao detector described by (41), which in this case is equivalent to (40), simplifies to [Kay and Sengupta 1986, C]

$$\ell_R = \frac{\left[\sum_{n=p+1}^{N}\left(\sum_{j=0}^{p}\hat{a}_j s_{n-j}\right)\hat{u}_n\Gamma(\hat{u}_n)\Big|_{\hat{u}_n=\sum_{j=0}^{p}\hat{a}_j x_{n-j}}\right]^2}{I_f(\hat{\sigma}^2)\sum_{n=p+1}^{N}\left(\sum_{j=0}^{p}\hat{a}_j s_{n-j}\right)^2} \tag{46}$$

where the dependence of I_f on $\hat{\sigma}^2$ is noted. ℓ_R was computed from the above expression for 500 different sets of data, each of length $N = 50$, for a given SNR. The theoretical value of the threshold γ' was used. The number of times ℓ_R exceeded γ', divided by 500 (the number of experiments) was regarded as the experimental value of the probability of detection. The MLE of the AR filter parameters required by (40) under $\mathcal{H}_0$ were replaced by the corresponding two-stage LS estimators.

The GLRT detector is given by (36*b*) with $p = 4$, $q = 1$. The MLE's were again replaced by the two-stage LS estimators.

Figure 3(*a*) plots the probabilities of detection of the Rao detector and the GLRT detector along with the theoretical performance vs. SNR for noise process I. Figure 3(*b*) plots the same for noise process II. The performance of the GLRT detector is very close to the theoretical performance. This confirms that a sample size of $N = 50$ is sufficient for the theoretical predictions of performance of the GLRT to be valid. Replacement of the MLEs by the computationally simple two-stage LS estimators appears to be a reasonable approximation. The performance of the Rao detector is equivalent to that of the GLRT detector for low SNR. However its performance deteriorates for stronger signals. The Rao detector essentially approximates a difference by a derivative [Rao 1973]. A large sample size and closeness of the two hypothses are crucial to this approximation. The failure of the Rao test at moderate SNR is an evidence of its inapplicability for short sample sizes.

VI. Conclusions

Two models of non-zero mean non-Gaussian AR processes were considered. The Cramer-Rao bounds for the parameters of each model were presented. A mismatch of the signal spectrum and the noise PSD favors precise estimation of the signal parameters. The CR bounds for the estimators of the AR filter parameters do not depend on the signal in the case of model (2). Estimation of these parameters in model (1) may be at least as precise, while a properly designed signal can reduce the CR bounds. In each case, the CR bounds for thse two sets of parameters are less in the case of a non-Gaussian PDF compared to a Gaussian PDF.

A two-stage LS estimator was proposed for the estimation of the signal coefficients and the AR filter parameters. This is an extention of a technique proposed earlier for zero-mean processes and is derived as an approximation to the MLE. The two-stage LS estimator provides a closed form solution to a set of approximated likelihood equations. This technique allows any signal shape for model (1), while for model (2) only permissible signal components are damped or undamped sinusoids, exponentials and dc.

The two-stage LS estimator was then applied to the GLRT detection of signals in colored non-Gaussian noise. Two mixing models were considered and the GLRT detector was derived for each one of them. The Rao detector for these problems were also presented. Computer simulations indicated the success of the GLRT and the failure of the Rao detector for moderate SNR and short sample sizes.

APPENDIX

Derivation of the Cramer Rao Bounds

The first step to determine the CR bounds is the evaluation of the information matrix for the parameter vector $\boldsymbol{\Phi}$ defined by (9) for process (1). One can proceed from (8) neglecting the second term on the right hand side.

$$\begin{aligned}
\mathbf{I}_{\boldsymbol{\Phi}} &\approx -E\left[\frac{\partial^2 \ln \mathbf{f}_c(\mathbf{x})}{\partial \boldsymbol{\Phi} \partial \boldsymbol{\Phi}^T}\right] \\
&= -\sum_{n=p+1}^{N} E\left[\frac{\partial^2 \ln f(u_n)}{\partial \boldsymbol{\Phi} \partial \boldsymbol{\Phi}^T}\right] \\
&= \sum_{n=p+1}^{N} E\left[\left(\frac{\partial \ln f(u_n)}{\partial \boldsymbol{\Phi}}\right)\left(\frac{\partial \ln f(u_n)}{\partial \boldsymbol{\Phi}}\right)^T\right] \\
&= \sum_{n=p+1}^{N} E\left[\left(\frac{\partial \ln \frac{1}{\sigma} g(\frac{u_n}{\sigma})}{\partial \boldsymbol{\Phi}}\right)\left(\frac{\partial \ln \frac{1}{\sigma} g(\frac{u_n}{\sigma})}{\partial \boldsymbol{\Phi}}\right)^T\right] \qquad (A.1)
\end{aligned}$$

where g is the PDF of the driving noise samples *scaled by* σ,

$$g(t) = \sigma f(\sigma t) \tag{A.2}$$

so that $g(t)$ always has unity variance. The partial derivatives are obtained as

$$\frac{\partial \ln\left(\frac{1}{\sigma} g\left(\frac{u_n}{\sigma}\right)\right)}{\partial \nu_i} = \left(\frac{g'\left(\frac{u_n}{\sigma}\right)}{\sigma g\left(\frac{u_n}{\sigma}\right)}\right)(-s_n^{(i)}), \quad i = 1, 2, \cdots, q \tag{A.3}$$

$$\frac{\partial \ln\left(\frac{1}{\sigma} g\left(\frac{u_n}{\sigma}\right)\right)}{\partial a_j} = \left(\frac{g'\left(\frac{u_n}{\sigma}\right)}{\sigma g\left(\frac{u_n}{\sigma}\right)}\right) x_{n-j}, \quad j = 1, 2, \cdots, p \tag{A.4}$$

$$\frac{\partial \ln\left(\frac{1}{\sigma} g\left(\frac{u_n}{\sigma}\right)\right)}{\partial \sigma^2} = -\frac{1}{2\sigma^2}\left[u_n\left(\frac{g'\left(\frac{u_n}{\sigma}\right)}{\sigma g\left(\frac{u_n}{\sigma}\right)}\right) + 1\right] \tag{A.5}$$

Note that

$$\left(\frac{g'\left(\frac{u_n}{\sigma}\right)}{\sigma g\left(\frac{u_n}{\sigma}\right)}\right) = \left(\frac{\sigma^2 f'(u_n)}{\sigma f(u_n)}\right) = \frac{f'(u_n)}{f(u_n)}$$

Hence

$$E\left[\left(\frac{g'\left(\frac{u_n}{\sigma}\right)}{\sigma g\left(\frac{u_n}{\sigma}\right)}\right)^2\right] = I_f \tag{A.6}$$

where

$$I_f = E\left[\left(\frac{\partial}{\partial u} \ln f(u)\right)^2\right]$$

Furthermore

$$E\left[\frac{g'\left(\frac{u_n}{\sigma}\right)}{\sigma g\left(\frac{u_n}{\sigma}\right)}\right] = 0 \tag{A.7}$$

$$\begin{aligned}
E\left[\frac{u_n g'\left(\frac{u_n}{\sigma}\right)}{\sigma g\left(\frac{u_n}{\sigma}\right)}\right] &= E\left[\frac{u_n f'(u_n)}{f(u_n)}\right] \\
&= \int_{-\infty}^{\infty} u_n f'(u_n) du_n \\
&= u_n f'(u_n) \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} f(u_n) du_n \\
&= -1
\end{aligned} \tag{A.8}$$

Entries of the information matrix can be readily obtained from $(A.3)$–$(A.5)$ using $(A.6)$–$(A.8)$ and a few other trivial identities such as the expectation of an odd function being zero.

$$E\left[\left(\frac{\partial \ln \mathbf{f}_c(\mathbf{x})}{\partial \nu_i}\right)\left(\frac{\partial \ln \mathbf{f}_c(\mathbf{x})}{\partial \nu_j}\right)\right] = \sum_{n=p+1}^{N} s_n^{(i)} s_n^{(j)} E\left[\left(\frac{g'\left(\frac{u_n}{\sigma}\right)}{\sigma g\left(\frac{u_n}{\sigma}\right)}\right)^2\right]$$
$$= I_f \mathbf{s}^{(i)T}\mathbf{s}^{(j)}$$

where $\mathbf{s}^{(k)} = [s_{p+1}^{(k)}\ s_{p+2}^{(k)}\ \cdots\ s_N^{(k)}]^T,\ k = 1, 2, \cdots, q$. Therefore

$$E\left[\left(\frac{\partial \ln \mathbf{f}_c(\mathbf{x})}{\partial \boldsymbol{\nu}}\right)\left(\frac{\partial \ln \mathbf{f}_c(\mathbf{x})}{\partial \boldsymbol{\nu}}\right)^T\right] = I_f \mathbf{S}^T\mathbf{S} \tag{A.9}$$

using the definition $\mathbf{S} = [\mathbf{s}^{(1)}\ \mathbf{s}^{(2)}\ \cdots\ \mathbf{s}^{(q)}]$.

$$E\left[\left(\frac{\partial \ln \mathbf{f}_c(\mathbf{x})}{\partial \nu_i}\right)\left(\frac{\partial \ln \mathbf{f}_c(\mathbf{x})}{\partial a_j}\right)\right] = -I_f \sum_{n=p+1}^{N} s_n^{(i)} m_{n-j} = -I_f \mathbf{s}^{(i)T}\mathbf{m}_j$$

where

$$m_n = E[x_{n-j}]$$
$$\mathbf{m}_j = [m_{p+1-j}\ m_{p+2-j}\ \cdots\ m_{N-j}]^T, \qquad j = 1, 2, \cdots, p$$

It follows that

$$E\left[\left(\frac{\partial \ln \mathbf{f}_c(\mathbf{x})}{\partial \boldsymbol{\nu}}\right)\left(\frac{\partial \ln \mathbf{f}_c(\mathbf{x})}{\partial \mathbf{a}}\right)^T\right] = -I_f \mathbf{S}^T\mathbf{M} \tag{A 10}$$

where $\mathbf{M} = [\mathbf{m}_1\ \mathbf{m}_2\ \cdots\ \mathbf{m}_p]$.

$$E\left[\left(\frac{\partial \ln \mathbf{f}_c(\mathbf{x})}{\partial \nu_i}\right)\left(\frac{\partial \ln \mathbf{f}_c(\mathbf{x})}{\partial \sigma^2}\right)\right]$$
$$= \sum_{n=p+1}^{N} (-s_n^{(i)})\left[E\left[u_n\left(\frac{g'\left(\frac{u_n}{\sigma}\right)}{\sigma g\left(\frac{u_n}{\sigma}\right)}\right)^2\right] + E\left[\frac{g'\left(\frac{u_n}{\sigma}\right)}{\sigma g\left(\frac{u_n}{\sigma}\right)}\right]\right]$$
$$= 0$$

Hence

$$E\left[\left(\frac{\partial \ln \mathbf{f}_c(\mathbf{x})}{\partial \boldsymbol{\nu}}\right)\left(\frac{\partial \ln \mathbf{f}_c(\mathbf{x})}{\partial \sigma^2}\right)\right] = 0_q \tag{A.11}$$

$$E\left[\left(\frac{\partial \ln \mathrm{f_c}(\mathbf{x})}{\partial a_i}\right)\left(\frac{\partial \ln \mathrm{f_c}(\mathbf{x})}{\partial a_j}\right)\right] = \sum_{n=p+1}^{N} E\left[\left(\frac{g'(\frac{u_n}{\sigma})}{\sigma g(\frac{u_n}{\sigma})}\right)^2 x_{n-i}x_{n-j}\right]$$

$$= I_f \sum_{n=p+1}^{N} E[x_{n-i}x_{n-j}]$$

$$= I_f \sum_{n=p+1}^{N} E[(x_{n-i} - m_{n-i})(x_{n-j} - m_{n-j}) + m_{n-i}m_{n-j}]$$

$$= I_f \left[(N-p)[\mathbf{C}]_{ij} + \mathbf{m}_i^T \mathbf{m}_j\right]$$

where C is the $p \times p$ covariance matrix of the data. Note that $x_n - m_n$ is a zero-mean wide sense stationary (WSS) process even though x_n has a time-varying mean. From the above equation it follows that

$$E\left[\left(\frac{\partial \ln \mathrm{f_c}(\mathbf{x})}{\partial \mathbf{a}}\right)\left(\frac{\partial \ln \mathrm{f_c}(\mathbf{x})}{\partial \mathbf{a}}\right)^T\right] = I_f[(N-p)\mathbf{C} + \mathbf{M}^T\mathbf{M}] \qquad (A.12)$$

$$E\left[\left(\frac{\partial \ln \mathrm{f_c}(\mathbf{x})}{\partial a_i}\right)\left(\frac{\partial \ln \mathrm{f_c}(\mathbf{x})}{\partial \sigma^2}\right)\right]$$

$$= -\frac{1}{2\sigma^2}\sum_{n=p+1}^{N} E\left[u_n\left(\frac{g'(\frac{u_n}{\sigma})}{\sigma g(\frac{u_n}{\sigma})}\right)^2 x_{n-i}\right] - \frac{1}{2\sigma^2}\sum_{n=p+1}^{N} E\left[\left(\frac{g'(\frac{u_n}{\sigma})}{\sigma g(\frac{u_n}{\sigma})}\right) x_{n-i}\right]$$

$$= -\frac{1}{2\sigma^2}\sum_{n=p+1}^{N} E\left[u_n\left(\frac{g'(\frac{u_n}{\sigma})}{\sigma g(\frac{u_n}{\sigma})}\right)^2\right] E[x_{n-i}] - \frac{1}{2\sigma^2}\sum_{n=p+1}^{N} E\left[\left(\frac{g'(\frac{u_n}{\sigma})}{\sigma g(\frac{u_n}{\sigma})}\right)\right] E[x_{n-i}]$$

$$=0$$

Therefore

$$E\left[\left(\frac{\partial \ln \mathrm{f_c}(\mathbf{x})}{\partial \mathbf{a}}\right)\left(\frac{\partial \ln \mathrm{f_c}(\mathbf{x})}{\partial \sigma^2}\right)^T\right] = \mathbf{0}_p \qquad (A.13)$$

$$E\left[\left(\frac{\partial \ln \mathrm{f_c}(\mathbf{x})}{\partial \sigma^2}\right)^2\right] = \frac{1}{4\sigma^4}\sum_{n=p+1}^{N} E\left[\left[u_n\left(\frac{g'(\frac{u_n}{\sigma})}{\sigma g(\frac{u_n}{\sigma})}\right) + 1\right]^2\right]$$

$$= \frac{1}{4\sigma^4}\sum_{n=p+1}^{N}\left[E\left[u_n^2\left(\frac{g'(\frac{u_n}{\sigma})}{\sigma g(\frac{u_n}{\sigma})}\right)^2\right] + 2E\left[u_n\left(\frac{g'(\frac{u_n}{\sigma})}{\sigma g(\frac{u_n}{\sigma})}\right)\right] + 1\right]$$

$$= (N-p)I_{\sigma^2} \qquad (A.14)$$

where

$$I_{\sigma^4} = \frac{1}{4\sigma^2}\left[E\left[\left(u\frac{\partial}{\partial u}\ln f\right)^2\right] - 1\right]$$

(10) follows directly from equations (A.9)–(A.14). The Cramer Rao bounds are obtained from the inverse of (10). (12c) results readily from the block diagonal property of the matrix. If $\mathbf{S}$ and $\mathbf{M}$ are scaled by σ and $\mathbf{C}$ is scaled by σ^2, then $\sigma^2 I_f$ can be pulled out from the four blocks of $\mathbf{I}_{\boldsymbol{\Phi}}$ corresponding to $\boldsymbol{\nu}$ and $\mathbf{a}$. One can then make use of the result

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{D} & \mathbf{C} \end{pmatrix}^{-1} = \begin{pmatrix} (\mathbf{A}-\mathbf{B}\mathbf{C}^{-1}\mathbf{D})^{-1} & -(\mathbf{A}-\mathbf{B}\mathbf{C}^{-1}\mathbf{D})^{-1}\mathbf{B}\mathbf{C}^{-1} \\ -(\mathbf{C}-\mathbf{D}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{D}\mathbf{A}^{-1} & (\mathbf{C}-\mathbf{D}\mathbf{A}^{-1}\mathbf{B})^{-1} \end{pmatrix}$$

As a consequence, the lower bound on $Cov(\hat{\boldsymbol{\nu}})$ is

$$\begin{aligned}
&\frac{1}{\sigma^2 I_f}\left[\mathbf{S}_n{}^T\mathbf{S}_n - \mathbf{S}_n{}^T\mathbf{M}_n[(N-p)\mathbf{C}_n + \mathbf{M}_n{}^T\mathbf{M}_n]^{-1}\mathbf{M}_n{}^T\mathbf{S}_n\right]^{-1} \\
&\quad= \frac{1}{\sigma^2 I_f}\left[\mathbf{S}_n{}^T\left[\mathbf{I} - \mathbf{M}_n[(N-p)\mathbf{C}_n + \mathbf{M}_n{}^T\mathbf{M}_n]^{-1}\mathbf{M}_n{}^T\right]\mathbf{S}_n\right]^{-1} \\
&\quad= \frac{1}{\sigma^2 I_f}\left[\mathbf{S}_n{}^T\left[\mathbf{I} + \mathbf{M}_n[-\mathbf{M}_n{}^T\mathbf{M}_n + (N-p)\mathbf{C}_n + \mathbf{M}_n{}^T\mathbf{M}_n]^{-1}\mathbf{M}_n{}^T\right]\mathbf{S}_n\right]^{-1} \\
&\quad= \frac{1}{\sigma^2 I_f}\left[\mathbf{S}_n{}^T\left[\mathbf{I} + \mathbf{M}_n[(N-p)\mathbf{C}_n]^{-1}\mathbf{M}_n{}^T\right]^{-1}\mathbf{S}_n\right]^{-1} \qquad (A.15)
\end{aligned}$$

using the identity

$$[\mathbf{A}+\mathbf{B}\mathbf{C}\mathbf{D}]^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}[\mathbf{D}\mathbf{A}^{-1}\mathbf{B} + \mathbf{C}^{-1}]^{-1}\mathbf{D}\mathbf{A}^{-1},$$

(A.15) is identical with (12a). Finally, using the inversion formula for block matrices once again, the lower bound on $Cov(\hat{\mathbf{a}})$ is

$$\begin{aligned}
&\frac{1}{\sigma^2 I_f}\left[(N-p)\mathbf{C}_n + \mathbf{M}_n{}^T\mathbf{M}_n - \mathbf{M}_n{}^T\mathbf{S}_n(\mathbf{S}_n{}^T\mathbf{S}_n)^{-1}\mathbf{S}_n{}^T\mathbf{M}_n\right]^{-1} \\
&\quad= \frac{1}{\sigma^2 I_f}\left[(N-p)\mathbf{C}_n + \mathbf{M}_n{}^T\left[\mathbf{I} - \mathbf{S}_n{}^T(\mathbf{S}_n{}^T\mathbf{S}_n)^{-1}\mathbf{S}_n\right]\mathbf{M}_n\right]^{-1} \qquad (A.16)
\end{aligned}$$

which is the same as (12b).

References

1. T.W. Anderson, *The Statistical Analysis of Time Series*, Chapter 5, New York: John Wiley, 1971.
2. G.E.P. Box and G.J. Jenkins, *Time Series Analysis: Forecasting and Control*, Chapter 7, San Francisco: Holden-Day, 1970.
3. P. Eykhoff, *System Identification*, Chapter 6, New York: John Wiley, 1974.
4. P.J. Huber, Robust Statistics, Chapter 4, New York: John Wiley, 1981.
5. S.M. Kay and D. Sengupta, "Simple and Efficient Estimation of Parameters for Non-Gaussian Autoregressive Processes", submitted for review to IEEE Trans. on Acoustics, Speech and Signal Processing, 1986.
6. S.M. Kay and D. Sengupta, "Detection in Incompletely Characterized Colored Non-Gaussian Noise via Parametric Modeling", submitted for review to IEEE Trans. on Acoustics, Speech and Signal Processing, 1986.
7. S.M. Kay and D. Sengupta, "Statistically/Computationally Efficient Detection in Incompletely Characterized Colored Non-Gaussian Noise via Parametric Modeling", submitted for review to IEEE Trans. on Acoustics, Speech and Signal Processing, 1986.
8. S.M. Kay and D. Sengupta, "Statistically/Computationally Efficient Estimation of Non-Gaussian AR Processes", presented at the International Conference on ASSP, Dallas, TX, April 6-9, 1987.
9. Sir S.M. Kendall and A. Stuart, *The Advanced Theory of Statistics Vol. II*, Chapters 18–19, New York: Macmillan Publishing, 1979.
10. C.H. Lee, "Robust Linear Prediction for Speech Analysis", presented at the International Conference on ASSP, Dallas, TX, April 6-9, 1987.
11. E.L. Lehmann, *Testing Statistical Hypotheses*, New York: John Wiley, 1959.
12. R.D. Martin, "Robust Estimation for Time Series Autoregressions" in *Robustness in Statistics*, edited by R.L. Launer and G.N. Wilkinson, New York: Academic, 1979.
13. R.D. Martin, "The Cramer-Rao Bound and Robust M-Estimates for Time Series Autoregressions", Biometrica, pp. 437–442, Vol. 69, No. 2, 1982.
14. C.R. Rao, *Linear Statistical Inference and its Applications*, Chapters 5–6, New York: John Wiley, 1973.

15. D. Sengupta and S.M. Kay, "Efficient Estimation of Parameters for Non-Gaussian Autoregressive Processes", submitted for review to IEEE Trans. on Acoustics, Speech and Signal Processing, 1986.

Table A: Parameters of the AR processes used for simulation

	a_1	a_2	a_3	a_4	poles
I	-1.352	1.338	-0.662	0.240	$0.7\exp[\pm j2\pi(0.12)]$ $0.7\exp[\pm j2\pi(0.21)]$
II	-2.760	3.809	-2.654	0.924	$0.98\exp[\pm j2\pi(0.11)]$ $0.98\exp[\pm j2\pi(0.14)]$

Table B: Performance of the unweighted LS Estimator, $N = 500$

		True value	Sample mean	Bias^2	Sample variance	Cramer-Rao bound
I	a_1	-1.3520	-1.3504	2.682×10^{-6}	1.9277×10^{-3}	2.0697×10^{-4}
	a_2	1.338	1.3377	8.862×10^{-7}	4.6654×10^{-3}	5.1217×10^{-4}
	a_3	-0.6620	-0.6621	3.434×10^{-9}	4.8690×10^{-3}	5.1217×10^{-4}
	a_4	0.2400	0.2425	6.081×10^{-6}	1.7834×10^{-3}	2.0697×10^{-4}
	μ	0.1000	0.0997	6.625×10^{-8}	1.1229×10^{-3}	1.1353×10^{-4}
	ν	0.4592	0.4578	1.899×10^{-6}	2.3824×10^{-2}	2.3959×10^{-3}
	σ^2	10.9000	10.8280	5.186×10^{-3}	6.0815	0.6323
II	a_1	-2.7600	-2.7544	3.134×10^{-5}	3.1395×10^{-4}	3.2114×10^{-5}
	a_2	3.8090	3.7944	2.124×10^{-4}	1.6867×10^{-3}	1.5815×10^{-4}
	a_3	-2.6540	-2.6394	2.232×10^{-4}	1.7395×10^{-3}	1.5815×10^{-4}
	a_4	0.9240	0.9174	4.391×10^{-5}	3.7182×10^{-4}	3.2114×10^{-5}
	μ	0.1000	0.0999	1.323×10^{-8}	1.9070×10^{-4}	1.9266×10^{-5}
	ν	1.1147	1.1092	3.052×10^{-5}	2.3450×10^{-2}	2.3939×10^{-3}
	σ^2	10.9000	10.8245	5.703×10^{-3}	6.0833	0.6323

Table C: Performance of the Two-stage LS Estimator, $N = 500$

		True value	Sample mean	Bias^2	Sample variance	Cramer-Rao bound
I	a_1	−1.3520	−1.3512	6.379×10^{-7}	2.7939×10^{-4}	2.0697×10^{-4}
	a_2	1.338	1.3369	1.276×10^{-6}	6.9665×10^{-4}	5.1217×10^{-4}
	a_3	−0.6620	−0.6616	1.987×10^{-7}	6.8015×10^{-4}	5.1217×10^{-4}
	a_4	0.2400	0.2401	3.283×10^{-9}	2.4486×10^{-4}	2.0697×10^{-4}
	μ	0.1000	0.0996	1.495×10^{-7}	1.2365×10^{-4}	1.1353×10^{-4}
	ν	0.4592	0.4572	4.131×10^{-6}	2.6284×10^{-3}	2.3959×10^{-3}
	σ^2	10.9000	10.9063	3.987×10^{-5}	6.1731	0.6323
II	a_1	−2.7600	−2.7588	1.403×10^{-6}	4.1403×10^{-5}	3.2114×10^{-5}
	a_2	3.8090	3.8058	1.040×10^{-5}	2.1812×10^{-4}	1.5815×10^{-4}
	a_3	−2.6540	−2.6506	1.151×10^{-5}	2.2206×10^{-4}	1.5815×10^{-4}
	a_4	0.9240	0.9224	2.410×10^{-6}	4.6890×10^{-5}	3.2114×10^{-5}
	μ	0.1000	0.0998	2.766×10^{-8}	2.1106×10^{-5}	1.9266×10^{-5}
	ν	1.1147	1.1119	7.776×10^{-6}	2.6332×10^{-3}	2.3939×10^{-3}
	σ^2	10.9000	10.9076	5.894×10^{-5}	6.1666	0.6323

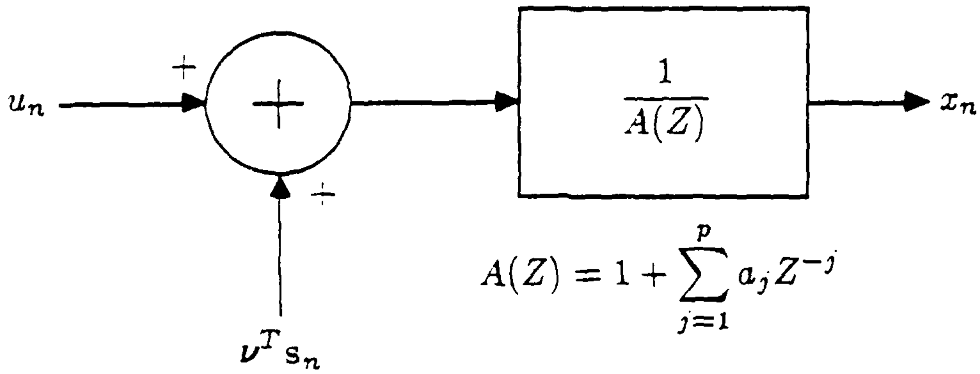

Figure 1(a): AR model for process (1)

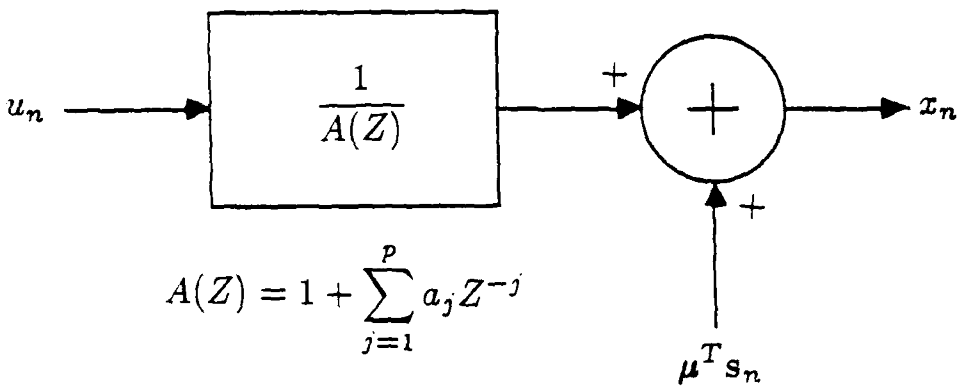

Figure 1(b): AR model for process (2)

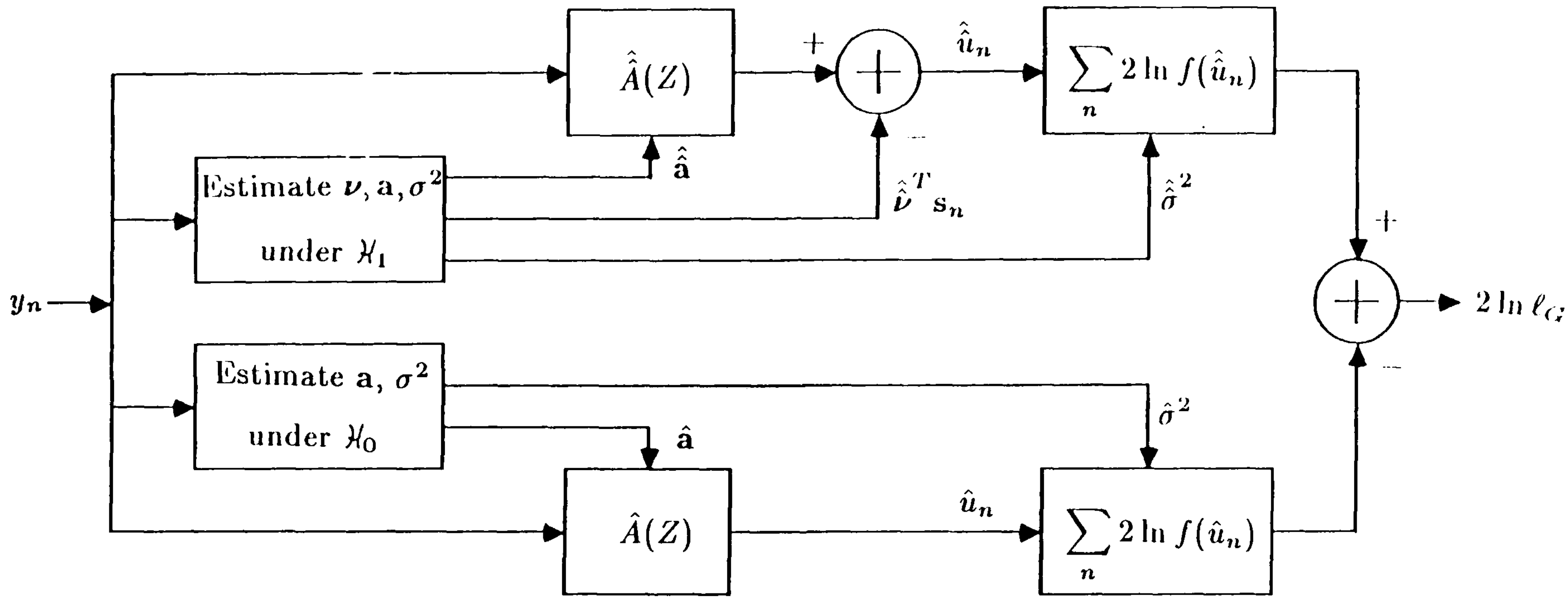

Figure 2(a): GLRT detector for problem (27)

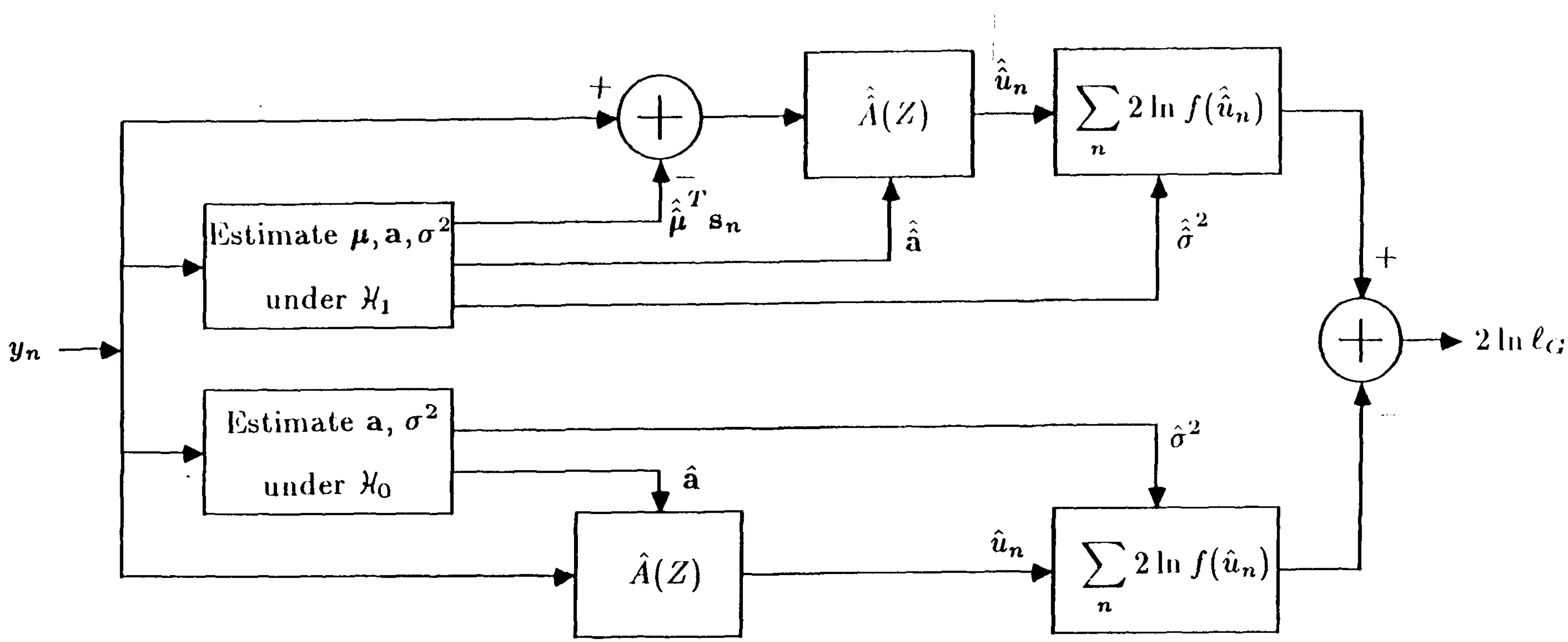

Figure 2(b): GLRT detector for problem (28)

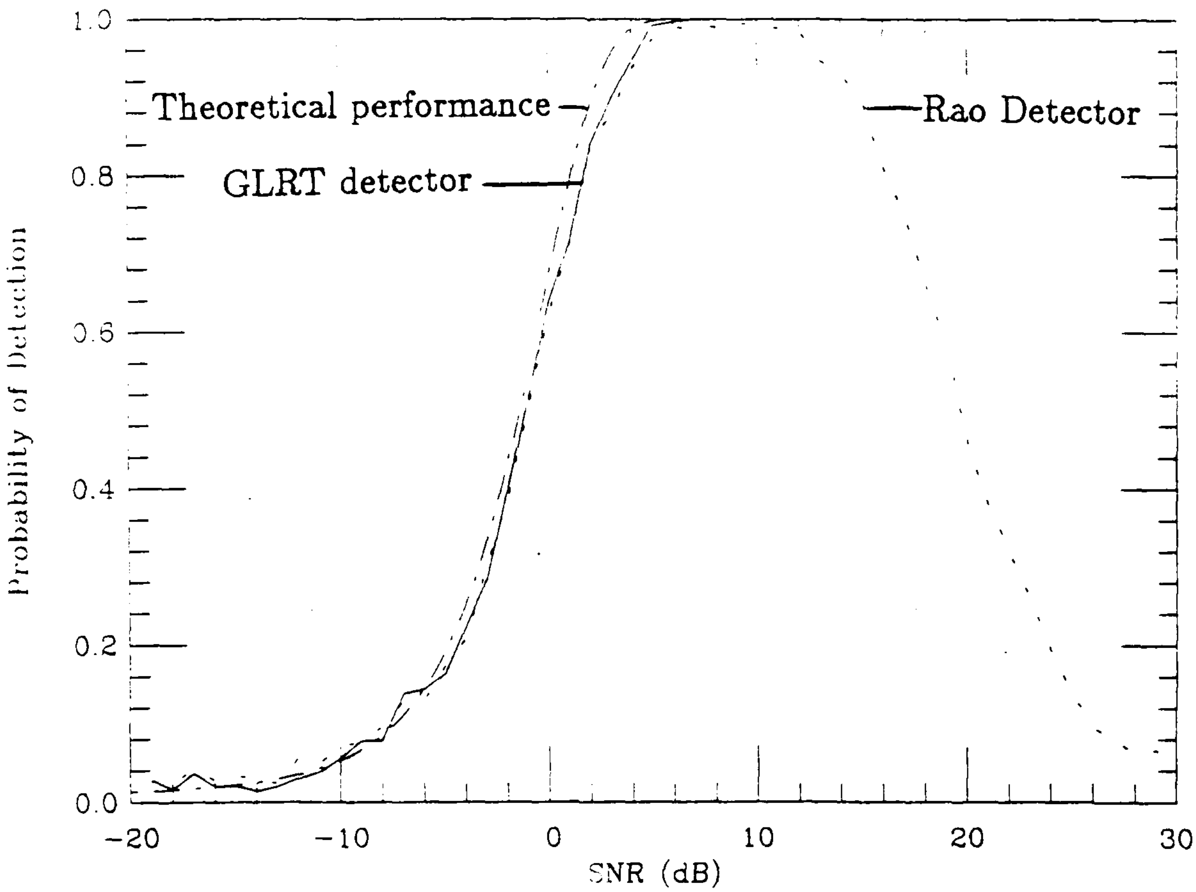

Figure 3(a): Performance of the GLRT and Rao detectors, process I

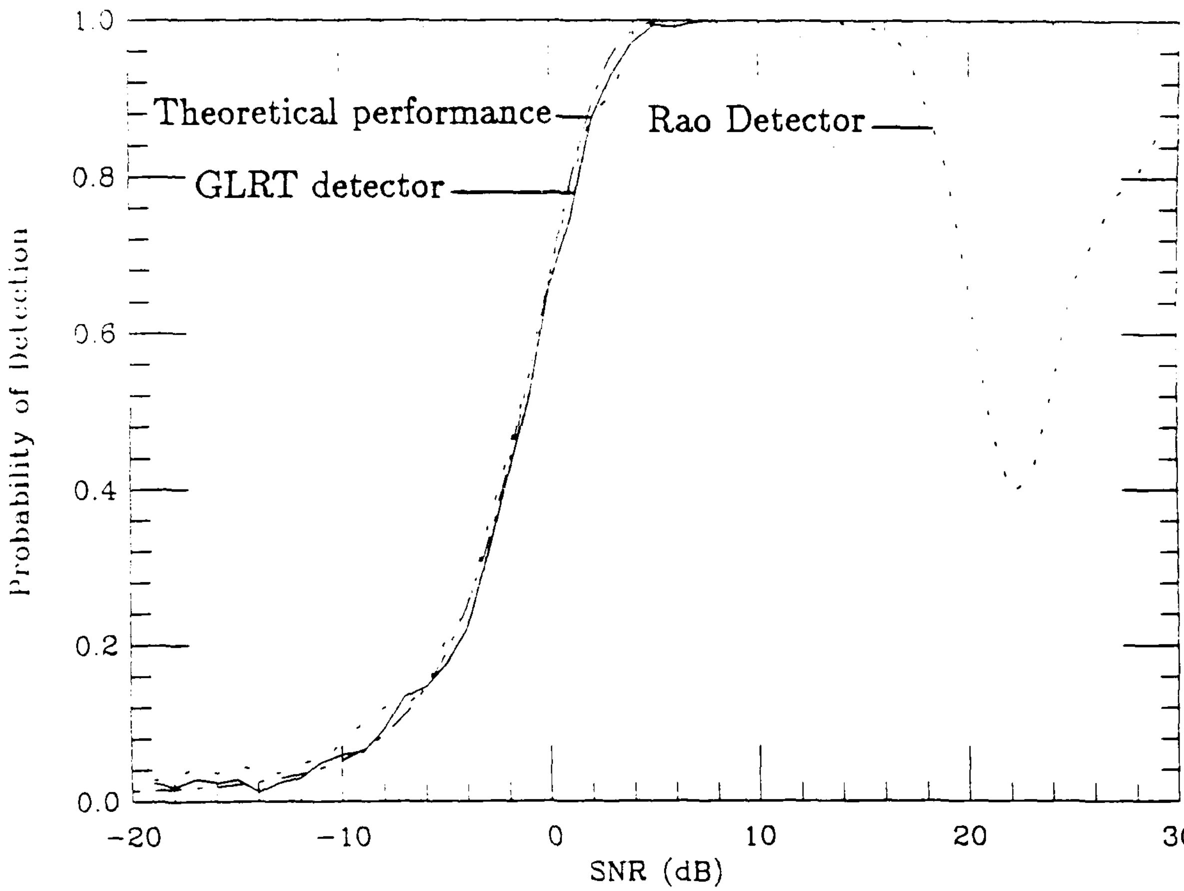

Figure 3(b): Performance of the GLRT and Rao detectors, process II